Requirements for an NLM Digital Repository: Report and Recommendations

Prepared by the
NLM Digital Repository Working Group

Submitted January 31, 2007
Revised final version March 16, 2007

Introduction

In order to fulfill the Library's mandate to collect, preserve and make accessible the scholarly and professional literature in the biomedical sciences, irrespective of format, it is essential that the Library develop the robust infrastructure needed to manage a large amount of material in a variety of digital formats. A number of Library Operations program areas are in need of such a digital repository to support their existing digital collections and to expand the ability to manage a growing amount of digitized and born-digital resources. The History of Medicine Division has created dozens of digital collections which require long-term management and preservation. Collection development and acquisitions staff are seeing an increasing availability of born-digital materials that NLM needs to add to its collection. NLM's preservation program has embraced digitization as a preservation method to replace microfilming.

In April, 2006, the Acting Associate Director for Library Operations approved the creation of a working group charged with developing functional specifications for an NLM Digital Repository and identifying policy and management issues related to the creation, design and maintenance of the repository. Working Group members were: Diane Boehr, Margaret Byrnes, Walter Cybulski, John Doyle, Laurie Duquette, David Gillikin, Jenny Heiland, Kuan-Tsae Huang, Felix Kong, Dianne McCutcheon (chair), Michael North, and John Rees. Lillian Kozuma worked with the group to prepare the inventory of existing and planned digital projects at NLM.

By identifying high level functional requirements and policy considerations, the NLM Digital Repository Working Group (DRWG) endeavored to outline an infrastructure and bring a standards-based approach to the management, preservation and access of NLM's existing and future digital resources.

The DRWG recommends the following as key next steps:

- LO and OCCS move forward to put in place a repository for the preservation of digital content not covered by PubMedCentral and the NIH CIT Videocast project. The NLM digital repository development should not duplicate these existing efforts and instead should focus on digital materials not covered by these two systems.
- Establish a Digital Projects Technical Group to coordinate the development of a digital repository and provide a technical review of LO digital projects.
- Evaluate commercial and open source software identified by the DRWG to identify a commercial system or open source software (or components of both) to be used as an NLM digital repository.

See the sections below for further details.

Functional Requirements Development (includes specifications for data formats and metadata requirements)
The DRWG developed high level functional statements following the Reference Model for an Open Archival Information System (OAIS)[1]. The DRWG followed the model since it is a common language in the digital repository environment, although the group recognized that it presented some challenges because the model intermingles functionality and policy.

Team members formed subgroups based on expertise and interest to draft the requirements for the various components of the OAIS model for review and discussion by the entire DRWG.

- Ingest (and Ingest Preparation): Laurie Duquette and Dianne McCutcheon
- Data Management, Preservation Planning and Archival Storage: Margaret Byrnes, Walter Cybulski, Felix Kong
- Administration: Margaret Byrnes, Walter Cybulski, Laurie Duquette, Felix Kong
- Access: David Gillikin, Jenny Heiland, John Doyle, Michael North
- Metadata: Diane Boehr, John Rees, Walter Cybulski

In order to create a common base of knowledge and understanding on key topics not familiar to the entire group, individuals volunteered to prepare briefings for the group. These presentations were very valuable in identifying important issues to be considered in the functional specifications.

- METS (Metadata Encoding and Transmission Standard): John Rees
- Descriptive Metadata Schemes: Diane Boehr
- Versions: Walter Cybulski

The group reviewed requirements documents created by GPO for the Future Digital System [2] and by the National Library of New Zealand for the National Digital Heritage Archive Programme [3]. These documents were helpful in terms of concepts to be included in the requirements, but the group chose to write the requirements statements from an NLM perspective. The NLM functional requirements document does follow the

---

[1] Consultative Committee for Space Data Systems (2002). "Reference Model for an Open Archival Information System (OAIS)". CCSDS 650.0-R-1 – Blue Book. Available at: http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf
[2] Government Printing Office, *Requirements Document (RD V1.0) for the Future Digital System (FDSys), Requirements Document (RD V2.0) for the Future Digital System (FDSys)* and *Concept of Operations (CONOPSV2.0) for the Future Digital System (FDSys)*
[3] Government Printing Office, *Requirements Document (RD V1.0) for the Future Digital System (FDSys), Requirements Document (RD V2.0) for the Future Digital System (FDSys)* and *Concept of Operations (CONOPSV2.0) for the Future Digital System (FDSys)*

approach taken by the New Zealand document of listing requirements by the OAIS model components.

Software Review
During the life of the DRWG, the NIH Library began investigation of software for use as an institutional repository. Margaret Byrnes, Laurie Duquette and David Gillikin were able to attend demonstrations of one or more of the software packages being considered by the NIH Library.   In addition, John Rees provided information on open source software which he had been exploring as part of his work in HMD. Due to this fortuitous opportunity to get a broad overview of existing commercial and open source software, the group was able to include recommendations in this report on selected software for further investigation by NLM for possible use as a digital repository.

A Note on the Process
"*Cultivating awareness is part of the process.*"- Walter Cybulski, October 24, 2006

The educational benefit of this effort to the group members should be mentioned. Building digital collections is a significantly different activity from building an analog collection with new technical challenges, changing roles and responsibilities for staff and creation of new relationships across the organizational structure. This process offered an opportunity for the group members to share the special expertise and experience of each participant, leading to a greater overall understanding of the requirements and challenges facing NLM in building a digital repository. The development of the requirements and recommendations resulted in creating a core group of LO staff representing all Divisions who gained knowledge of metadata standards, digital file and format types, and software developments for digital repositories.

Deliverables

Inventory of existing and planned digital projects at NLM
The inventory of existing and planned digital projects at NLM has been updated as of the end of January, 2007 and is submitted as a separate document along with this report. [*Internal working document; not available on the NLM website*]

Functional Requirements Development (includes specifications for data formats and metadata requirements)
The functional requirements for the NLM Digital Repository are submitted as a separate document, *Policies and Functional Requirements Specification for the National Library of Medicine Digital Repository Version 1*.

The NLM Digital Repository is envisioned as one or more electronic storage systems within which digitized and born-digital objects created, acquired, harvested or purchased by NLM reside. The repository has the ability to accept, store, preserve and provide access to all types of digital objects. Functionality includes:

- ingest and management of content as well as the descriptive, administrative and structural metadata associated with stored objects;
- preservation of objects in approved formats;
- controls to insure only permitted access to objects; and
- migration to new formats to insure objects do not become obsolete.

The DRWG chose to use the term "digital repository" instead of "digital archive". The literature is inconsistent in the definition of "digital archive", sometimes including access along with ingest, storage, preservation and migration functions and sometimes not. The more recent literature also indicates a preference for the term "repository" over "archive".

Policy Issues
The DRWG identified policy issues related to the digital repository and presented them in the functional requirements document under the following categories: 1) Relevant Assumptions; 2) Existing Policies; 3) Policies to be Reviewed and 4) Policies to be Developed.

Relevant assumptions are statements that the DRWG identified as related to the Digital Repository and believe to be accepted as true. The assumptions cover design requirements, scope and content of the repository and management issues. Prior to embarking on further digital repository development, NLM management needs to concur with these assumptions.

Existing policies are established NLM policies on collection development, digitization selection criteria, permanence levels and computer security that should be used to guide content selection and access controls for the digital repository.

Policies to be reviewed include the adoption of the TEI DTD in addition to the NLM Book and NLM Historical Book DTDs for encoding certain text materials. A detailed proposal is included in Attachment C of this report.

The NLM Metadata scheme will also need review. NLM has published an approved metadata schema based on Dublin Core (http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html). The needs of the Repository may require a new schema, more congruent with the existing NLMCommon DTD and the NLMCatalogRecord DTD.

A key area identified that will need policy development is migration strategies and methodologies. Digital preservation is a combination of storing objects in formats that can be migrated, recording appropriate metadata to be able to manage the objects, ongoing monitoring for bit and media degradation and technology obsolescence, and performing migration as needed. The preservation planning section of the functional requirements outlines the policy areas that need to be developed including monitoring changes in technology, evaluating the content in the archive, participating in standards development and developing migration plans.

Other areas which will need policy development relate to ingest into the repository, including which staff will be authorized to negotiate submission agreements, minimum requirements for content submitted by producers outside of NLM and maximum embargo periods.

<div style="background-color:#d9d9d9; border:1px solid black; padding:4px;">

Design Issues

</div>

Two major "design" issues arose during the deliberations of the DRWG: 1) a single digital repository system vs. multiple systems and 2) use of commercial software vs. open source software. Both sides of these issues have their proponents - within the DRWG, as well as in the larger digital repository community. These are extremely important areas which will need to be resolved by NLM management in order to move forward with decisions on the digital repository. Summarized below are the key ideas discussed by the DRWG.

- Single system vs. multiple individualized systems
  NLM currently has processes in place for the ingest, management, storage, archiving and access of digital material for the following acquisition and ingest streams: electronic journals deposited in PubMedCentral (PMC); digitized back files of PMC journals; and CIT Videocasts. There is a need to put in place a reliable repository for the preservation of digital content not covered by PMC and the videocast project to ensure ongoing access.

  LO has reached a point when migration from isolated digital projects to a more integrated digital repository environment is needed. The DRWG recognizes that while the ideal goal would be to have a single system for access and preservation of digital objects, we also realize that this is not feasible immediately. However, having multiple systems creates a problem for keeping data in sync, is more difficult for OCCS to support, and raises issues of staffing capability.

  The goal for the NLM Digital Repository should be to move toward fewer solutions, rather than continuing the proliferation of isolated digital projects that are supported by different retrieval systems, and to bring them under a controlled environment in the move toward an integrated digital collection. We need to look for systems that can accommodate most of our needs rather than systems than meet needs for individual projects. At the same time, we need to be flexible so that exceptions can be made if projects are proposed that cannot be accommodated by the main system.

- Commercial software and/or open source software
  The DRWG's software evaluation included review of commercial as well as open source repository software. The review focused primarily on functionality and scalability of the software. However, there are distinct tradeoffs in selecting a commercial repository solution compared with using open source software (or developing an in-house solution) including: cost (to purchase or to develop),

ongoing development overhead, support overhead and flexibility. As part of a more detailed evaluation of the recommended software to be done in the next phase, NLM needs to include analysis of the development and support needs of the two approaches. If there are no commercial systems that offer all the needed functionality, the NLM repository may need to be a combination of commercial and open source software.

Recommendations for the Next Phase

1. Establish a Digital Projects Technical Group

   a. The Technical Group should be composed of members from Library Operations and OCCS.
   b. The Technical Group will do a technical review of all proposed digitization projects (as well as born-digital acquisitions) to determine what special needs are involved and whether the existing system can accommodate them.
   c. The Technical Group will coordinate efforts with the Digital Projects Selection Group and ensure that preservation, minimum technical and metadata requirements, access approach, etc. have been adequately addressed during the planning phases of every new digital project. Individual group members would have specific responsibility for approving one or more parts of the plan, based on their area of expertise such as metadata formats, digitization standards, etc.
   d. The Technical Group will develop a checklist and digital projects database to track proposed and approved digital project plans. Signoff form/Checklist evaluation criteria should include:
      i. Access plan – method, degree of access based on intellectual property considerations
      ii. Metadata plan – Source of descriptive metadata, etc.
      iii. Digitization standards used
      iv. Preservation plan
      v. If separate system is recommended, provide business case for its use and payoff for the organization in not using existing system.

2. Evaluate commercial systems and open source software identified by the DRWG to identify a commercial system or open source software (or both) for use as an NLM digital repository.

   a. The Digital Projects Technical Group should be tasked to take the functional specifications created by the DRWG and do in-depth analysis of the commercial systems and open source software identified by the DRWG. The Digital Projects Technical Group may want to investigate additional software that may be available.

b. The systems/software evaluation will include analyses of how well each system/software meets the functionality of the OAIS model – ingest, archival storage, data management, administration, preservation planning, and access; and the risks involved in using commercial systems and/or open source software.

See Appendix A for further details on the evaluation and Appendix B for the spreadsheet of commercial systems and open source software.

3. Adopt the TEI DTD in addition to the NLM Book and NLM Historical Book DTDs for encoding certain text materials. See Attachment C for background and justification.

4. Participate in additional groups working on standards and issues related to the digital library environment to keep abreast of ongoing developments and have input into new standards. The DRWG specifically recommends that NLM consider joining the Digital Library Federation, either as a strategic member or as an allied member, so that NLM staff can participate on working groups and as part of the DLF Forum. http://www.diglib.org/about/dlfmission.htm

5. Monitor technologies and modify the functional requirements as appropriate.

<p style="text-align:center">Appendix A</p>

NLM Digital Repository Systems/Software Evaluation

Goals:

The goals of the NLM Digital Repository systems/software evaluation encompass the following: a determination on whether a commercial system or open source software or a combination of both is the most feasible approach for the NLM digital repository; how well these systems meet the functionality of the OAIS model – ingest, archival storage, data management, administration, preservation planning, and access; and an analysis of the risks of commercial systems and/or open source software.

Methodology:

The evaluation will be a two part process of evaluating commercial systems and open source software (see attached document that details each process). The commercial system vendors and open source software identified by the NLM Digital Repository Group will be used in the evaluation (see the attached spreadsheet), along with additional software or systems that may be identified by the Digital Projects Technical Group.

The evaluation will cover the following areas:

- Ingest methods
- Import/export capabilities
- Formats/file types supported
- Metadata schemes used
- User interface
- Search capabilities – full text; federated; browse
- Rights management
- Image presentation
- Standards compliance
- Scalability, interoperability and system performance
- Security features
- Preservation functions

Assumptions:

- o A checklist of functionality gleaned from the NLM Digital Repository functional requirements specification will be used for evaluation
- o The same set of diverse objects and metadata will be used in the evaluation
- o The same group of testers will participate in the evaluation

Questions:

Can a simultaneous test be accomplished at the same time of commercial systems and open source software?

Commercial System Evaluation Process

Assumptions:
- A minimum of a 90 day trial should be conducted
- Reviews of vendor customer sites on the internet will be performed
- Phone calls and/or site visits will be conducted to vendor customers
- Costs may be associated with procuring trials of commercial systems. NLM may need to provide funding in order to perform the evaluations.
- Access to test systems at commercial vendor sites through the NLM firewall will be granted by OCCS if needed
- Installation of vendor software in the NLM environment will be allowed by OCCS if needed

Questions:
What is involved and who should be involved in implementing a commercial system?
What should be NLM's approach to using a commercial product?
Will a simultaneous test be conducted on each commercial system?
Can this product be used with another product?

Open Source Software Evaluation Process

Assumptions:
- The testing period for open source software may vary depending on the amount of configuration and set-up needed to test at NLM.
- Reviews of existing open source sites on the internet will be performed
- Existing open source literature reviews will be evaluated
- Phone calls and/or site visits to existing open source users will be performed
- Installation of open source software in the NLM environment will be allowed by OCCS if needed

Questions:
Open source – who will do the programming, will it be OCCS or LHC staff or both?
What is involved and who should be involved in implementing open source software?
What should be NLM's approach to using open source software?
Will there be a simultaneous test of each open source software?
Can this product be used with another product?

Appendix B
Software and Systems to Be Considered

Software Name (Vendor/OpenSource) X = Yes, software provides ? = unknown	ArchivalWare (PTFS)	ContentDM (OCLC)	DigiTool (ExLibris)	DSpace (open source)	Fedora (open source)	Greenstone (open source)	Keystone DLS (IndexData opensource)	SPER (open source)
Storage For								
Born Digital	?	X	X	X	X	X	?	X
Digitized	X	X	X	X	X	X	X	X
Web Harvested	?	?	?			X	X	X
Features								
Ingest Method (How is material ingested?)	batch and ?	templates or batch	submission forms	single, batch	batch utility	batch, single, templates via scripts	data entry, automated scripts, crawls	batch/single
Formats (What formats are supported?)	JPEG 2000, TIFF, DjVu, PDF, XML	JPEG 2000, TIFF, PDF, XML,	TIFF, PDF, XML, JPEG2000, XLS, CSV, MARC	TIFF, PDF, GIF, JPEG, XML, MARC, HTML, MPEG	TIFF, XML, JPEG, GIF, PDF	XML, TIFF, GIF, JPEG, MARC, PDF, HTML	XML, JPEG, PDF, HTML, GIF, XML	TIFF, JPEG2, JPEG, GIF, XML
File/Object Types (What file types are supported?)	audio, video, images, multipage text, engineering drawings, MS Office	audio, video, images, text, 6 sided objects	audio, video, multi-page text, images	audio, video, text, image,	text, image, audio, video	text, images, html, audio, video, MS Office	text, images, audio, video, MSOffice	images, text
Metadata Schemes (What metadata schema does it handle?)	DCPlus	DC, VRA2.0, METS, EAD, MODS	DC, METS, EAD, TEI, VRA2.0,	DC, METS	DC/QDC, METS, XML	METS, XML, DC	DC/QDC, METS, TEI, XML	any xml base
Metadata Editing (Can metadata be edited?)	X	X	X	X	X	X	X	X
Full Text Searching (Does it allow searching of full text and highlight results?)	X	X(highlighting if using OCR extension software)	X(some highlighting)	X(no highlighting)	X(no highlighting)	X	X (highlighting once doc opened)	X
"Federated" searching (Does it do cross collection searching?)	X	X	X	X	X	X	X	N/A
Browsing (Does it support browsing of collections?)	X	X	X	X	X	X	X	N/A
Image Presentation (What and how are images presented?)	Thumbnails, Full image Zooming	Thumbnails, Full image Zooming	Thumbnails, Full Image Zooming via Mr Sid/JPEG2000	link to full image	Thumbnails, Full Image Zooming	Thumbnails, Small/Full Image	Thumbnails	native format
Rights Management	?	X	X	X	X	X	X	N/A
Import/Export of Data (Single or Batch import/export allowed?)	?	X	X	X (METS export)	X	METS export; import to and export to Dspace	X; export is MARC	X (XML)
Scability (Can it scale to accommodate millions of objects?)	No limit (largest collection to be over 10 million)	No limit	No limit (largest collection 80,000 records)	Limited to 100,000 records	unlimited (largest collection around 1 million records)	unlimited(largest 11 million short docs)	Using DSpace1.x=200,000; using DSpace2.0=millions	
Standards Compliance (what standards does it comply with?)	Unicode ,JPEG2000, ?	JPEG2000, OpenURL, Z39.50, OAI-PMH	JPEG2000, OpenURL, Z39.50, OAI-PMH, PREMIS Support	OAI-PMH, SRU/SRW, SOAP, LDAP, Unicode	OAI-PMH, LDAP, SOAP, Unicode	OAI-PMH, Unicode, Z39.50	OAI-PMH; SOAP, Z39.50, SRU/SRW; Unicode	SOAP, METS, Z39.87
Existing Production Sites	http://www.ptfs.com/products_case.html	www.dimema.com/customers/index.html	http://www.exlibrisgroup.com/customers/index ool_publications.htm	http://wiki.dspace.org/index.php/Dsp aceInstances	http://www.fedora.info/co mmunity/	http://www.greenstone.or g/cgi-bin/library?e=p-en-home-utfZz-8&a=p&p=examples	see references at http://www.indexdata.dk/ keystone/	

Software Name (Vendor/OpenSource)	DigitalCommons (ProQuest/UMI)	DLXS (open source)	ENCompass/Curator (Endeavor)	Insight (Luna Imaging)	Internet Archive	PubMed Central (NLM)	Streetprint.org (open source)
X = Yes, software provides							
? = unknown							
Storage For							
Born Digital	X	X				X	
Digitized	X	X	X	X		X	X
Web Harvested					X	expected in the future	
Features							
Ingest Method (How is material ingested?)	template input, batch review to host server	single/batch xml	batch import of XML	batch data import of XML	hosted crawls or subscription crawls	batch data import	templates and uploader
Formats (What formats are supported?)	PDF, HTML	TIFF, PDF, GIF, JPEG2000, XML	TIFF, JPEG, PDF, XML, TEXT, MPEG, GIF, HTML, MARC	JPEG2000, XML, MARC	XML, PDF	XML, JPEG, PDF, TIFF, HTML	JPEG
File/Object Types (What file types are supported?)	text, and as attachments - Audio, Video, MS Office	text, image	audio, video, text, image, MS Office	images, audio, video, 3D images, multi-page text	html, text, images, audio, video, MS Office	html, text, images, audio, video, MS Office	images, audio, video, text, MS Office
Metadata Schemes (What metadata schema does it handle?)	HTML only	XML (does not natively handle METS; doesn't mean you can't XLINK out to another application)	DC/QDC, EAD, TEI, MODS, custom	XML, DC, VRA2.0, custom templates	DC	XML, NCBI DTD	DC
Metadata Editing (Can metadata be edited?)	X	X	X (also technical)	X	X	X	X
Full Text Searching (Does it allow searching of full text and highlight results?)	X (no highlighting)	X	X(if fulltext indexed)(no highlighting)		X	X (highlights if user access through My NCBI)	X (no highlighting)
"Federated" searching (Does it do cross collection searching?)	X	X	X		X	X (other Entrez databases)	X
Browsing (Does it support browsing of collections?)	X	X	X	X		X	X
Image Presentation (What and how are images presented?)	?	Thumbnails, Full Image Zooming (via Mr.Sid/JPEG2000)	Thumbnails	Thumbnails, Full Image Zooming via JPEG2000		Thumbnails, Full images	Thumbnail, Full Image
Rights Management	X	X	X	X		Some	X
Import/Export of Data (Single or Batch import/export allowed?)	X (export is in XML)	X	X	X		X	Import only ?
Scability (Can it scale to accommodate millions of objects?)	Based on disc space purchased	Limited (current largest collection is 3.5 million pages [12K volumes])	unlimited (largest collection at NLM 4.7 million records)	unlimited	subscription allows 10 million docs per year	Yes	?
Standards Compliance (what standards does it comply with?)	OAI, NLM DTD and DC DTD for export	JPEG2000, Unicode, OAI-PMH	Unicode, OpenURL, SRU/SRW, OAI-PMH	JPEG2000, OAI	OAI-PMH	NCBI DTD	OAI, SOAP
Reason for no further consideration:	Hosted service only	Scalability is a major concern	Long term feasibility of product is a concern now that Endeavor has been merged with ExLibris	Software for images only storage	OK to use for web harvesting but does not warrant use for other storage	Used for journals only	Only format supported is JPEG

Attachment C
Recommendation to use the TEI DTD for NLM Digital Repository

We recommend the TEI DTD in addition to the NLM Book and NLM Historical Book DTDs for encoding certain text materials. There are three major reasons: rationale for the DTDs; element availability; user constituency

Rationale:
The NLM Book and Historical Book DTDs are conceptually limited in that they assume the encoding is to be a book or similar codex format rather than simply objects bearing text. The NLM DTDs did not consider the need to encode formats such as letters, oral history transcripts, plays, broadsides, diplomas, lecture notes, etc., all of which can be found within NLM's collections. TEI was created for the textual analysis of any format and the unique characteristics each bears and can conceptually embrace and define all of the above formats. Moreover, the NLM DTDs were not created with the intent to preserve the intrinsic characteristics of a text-bearing object but instead were created to normalize the NCBI archive's content that is acquired from publishers and their myriad individual DTDs. TEI respects the need to preserve the complete text as an authentic object with all the flaws and idiosyncrasies the history of printing entails.

Element Availability:
The conceptual limitation of the NLM Book and Historical DTDs leads to the understandable lack of elements available for encoding texts in the documentary edition and linguistic analysis traditions. While the Historical Book DTD attempts to rectify some of these omissions, for example page break milestones, word normalization, and textual annotations, they are not properly defined or useful in the contexts we envision. Among others, TEI contains specific elements for editorial interventions such as corrections, regularizations, sic, identifying original vs. normalized versions, etc. TEI also allows for the specific encoding of deletions, omissions, and additions such as intralinear, supralinear, inline, left, right, gaps, unclear, responsible party, type and many others. Also, the NLM DTDs are designed to be constraining (necessary for normalization) whereas TEI is more accommodating in its elements and CDATA attributes to allow for any vagary that a text could contain and not previously defined by the DTD.

User Constituency:
The NLM DTDs were designed for E-books and medical textbooks primarily for the NLM Bookshelf, that is, the medico-scientific community. Another primary NLM audience is the humanities scholar, whether it be a historian, social scientist, political scientist or the like. The needs and expectations of this audience is quite different from the medico-scientific community. TEI was first conceived in 1987 (DTD first draft, 1990), primarily by the Linguisitic and English Literature community and has since accommodated the other humanities fields. Moreover, the longevity, breadth, and depth of TEI's adoption by these communities has created a universe of texts and a preponderance of encoding, analytic, and presentation tools not available with the NLM

DTDs. Lastly, TEI allows the sharing and interoperability of our texts with this large community that is not possible with the NLM DTDs.